Dispute as rival groups publish details of human genome

Gavin Yamey *BMJ*

The two groups that are sequencing the human genome published full sequence and mapping details of their first drafts this week, an event marked by acrimonious fighting between the rival scientists involved.

The Human Genome Project, an international collaboration between eight publicly funded university centres, published its sequence data in the journal *Nature*. Its data have been posted daily on the internet since the project began and are freely available. The project estimates that in the past two months researchers in the developing world have accessed its genome database over 300 000 times.

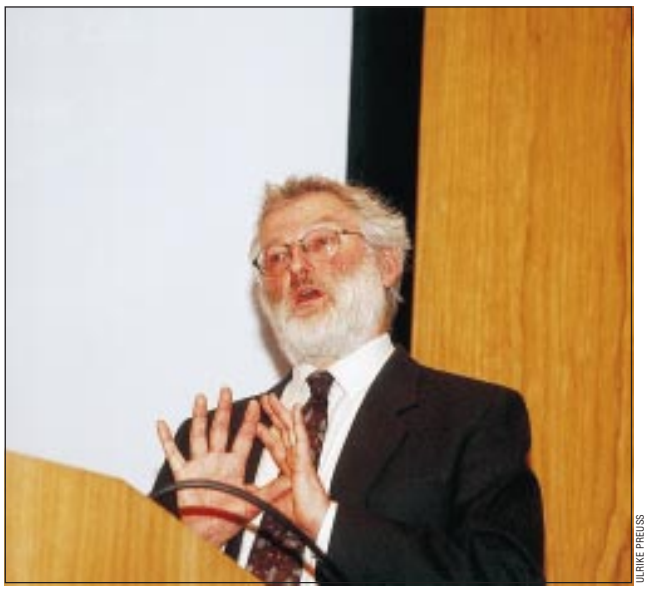
In contrast, the commercial enterprise Celera Genomics, which published its data in the journal *Science*, only allows paid subscribers to access its data. *Science* took the unprecedented step of agreeing to impose some restrictions on the use of the published Celera data. The Human Genome Project, angered by the way that these restrictions discriminate against scientists in poor countries, claimed that Celera had attracted fewer than 50 paying customers.

"Others want to charge the rest of the human race a fortune, but we are here to tell them that the human genome is not for sale," said Sir John Sulston, former director of the Sanger Centre in Cambridge, which sequenced a third of the public project's genome.

"Our international publicly owned data," he said, "are benefiting local communities." The sequence data, he explained, are being used by scientists in the developing world to study genetic variations in susceptibility to common fatal diseases, such as diarrhoea, malaria, and AIDS.

Craig Venter, Celera's founder, dismissed the accusations of profiteering as "a minor squabble between scientists," and Barbara Jasny and Donald Kennedy, the editors of *Science*, said that the rivalry "detracts from the awesome accomplishment jointly unveiled this week."

But while the *Science* editors talked of "a marriage between public funding and private entrepreneurship," the public project claimed that the relationship was far from equal, and that Celera's shotgun method for assembling sequence data relied on the public genome database.



ULRIKE PREUSS

Sir John Sulston: "We are here to tell them that the human genome is not for sale"

Richard Durbin, the deputy director and head of bioinformatics at the Sanger Centre, said: "They [Celera] have relied very heavily on the work we put in to put together our sequence. The message at the end is that, although they have added some new material, overall the results are remarkably comparable."

Despite the different sequencing techniques used by the two groups, both agreed that the human genome contains far fewer genes than was originally estimated when their first working drafts were announced last year (*BMJ* 2000;321:7).

The drafts suggested that there might be up to 115 000 genes, but the final number is between 30 000 and 40 000, making the human genome only twice as large as that of the fruit fly.

The surprisingly small size of the genome led to widespread media speculation that our environment must influence our development more than our genes do. But both groups rejected this notion, pointing out that our genes must interact in a myriad ways to drive human complexity, variation, and disease.

"The notion that one gene equals one disease," said Dr Venter, "or that one gene produces one key protein, is flying out the window."

Another key finding is that humans share many of their genes with simple organisms—half with nematodes and a fifth with yeast. Michael Dexter, director of the Wellcome Trust, which has committed £210m ($315m) to the public project, believes that this confirms Darwin's theory of evolution. "It's great to be getting the molecular correlates of what Darwin hypothesised 150 years ago," he said.

More than a third of the genome contains repetitive DNA sequences ("junk DNA"), much more than in any other genome sequenced to date. Scientists now believe that far from being useless, these DNA repeats have been crucial to the evolution of the human genome, mediating the creation of new genes. "The junk in the genome," said Richard Gallagher, *Nature*'s chief biology editor, "offers a window into our history. There are a huge number of stories to be uncovered."

Publication of the human genome ushers in a new era of post-genomic science, in which researchers will use the data to try to understand the causes of diseases and to develop new treatments. A paper published in *Nature*, for example, suggests that the draft sequence will help us to understand the biology of drug addiction by enabling us to identify "addiction vulnerability genes" (*Nature* 2001;409:834-5). Other papers point to how the sequence data could provide new treatments for sleep disorders and jet lag. □

Sequence data, plus commentaries and analyses, are available at www.nature.com/genomics/human and www.sciencemag.org/genome 2001